

Zadání 3. domácího úkolu

1 Pokyny k odevzdání

Řešení vypracujte v libovolném textovém editoru, na který jste zvyklí (např. Word atd.), a nahrajte do SIS → Studijní mezivýsledky. Pokud by byl v úloze nalezen nějaký nedostatek, který by potřeboval opravu, bude úloha vrácena k přepracování. Prosím, snažte se, aby zasílané řešení bylo co nejvíce správné a nebylo tak třeba úlohu vracet. Čas na vypracování není omezen, ale úkol odevzdejte s dostatečným předstihem před plánovaným termínem zkoušky.

2 Pokyny k vypracování

Řešení bohatě okomentujte celými větami, aby bylo patrné, že použitým metodám rozumíte. Všechny své odpovědi řádně zdůvodněte a interpretujte. Do vašeho dokumentu vložte veškeré R-příkazy, které jste při vypracování použili (buď příkazy nakopírujte na příslušná místa do textu, nebo skript vložte na konec dokumentu). Nezasílejte prosím vypracování ve formě R-kového skriptu.

3 Data

V úloze využijeme datovou tabulku `WhiteWine.csv`, která obsahuje údaje o chemických vlastnostech bílé varianty portugalského vína „Vinho Verde“. Jedná se součást většího datového souboru nasbíraného od května 2004 do února 2007, který je detailněji popsán v článku: P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis: Modeling wine preferences by data mining from physicochemical properties, *Decision Support Systems*, Elsevier, 47(4):547-553, 2009.

Data obsahují následující veličiny (originální popis):

`fixed acidity`: $g(\text{tartaric acid})/dm^3$

`volatile acidity`: $g(\text{acetic acid})/dm^3$

`citric acid`: g/dm^3

`residual sugar`: g/dm^3

`chlorides`: $g(\text{sodium chloride})/dm^3$

`free sulfur dioxide`: mg/dm^3

`total sulfur dioxide`: mg/dm^3

`density`: g/cm^3

`pH`

`sulphates`: $g(\text{potassium sulphate})/dm^3$

`alcohol`: vol. %

`quality`: number from 0 (very bad) to 10 (excellent)

`type`: white

Soubor se načte do R pomocí příkazu:

```
Wine <- read.csv2("WhiteWine.csv", header=TRUE, sep=",")
```

4 Zadání

Vaším úkolem bude prozkoumat závislost obsahu alkoholu na hustotě vína.

1. Spočtete Pearsonův korelační koeficient mezi obsahem alkoholu (veličina `alcohol`) a hustotou (veličina `density`) a hodnotu interpretujte.
2. Uvažujte lineární regresní model, ve kterém bude odezvou obsah alkoholu (`alcohol`) a regresorem bude hustota vína (`density`). Půjde tedy o model jednoduché regrese. Zapište teoretický tvar modelu včetně předpokladů.
3. Odhadněte všechny parametry modelu (tj. regresní koeficienty a směrodatnou odchylku náhodných chyb modelu). Zapište konkrétní tvar odhadnuté regresní přímky.
4. Je regresor `density` v modelu významný? Jak to poznáte? (Nezapomeňte uvést tvar příslušné nulové a alternativní hypotézy, p-hodnotu a váš závěr.)
5. Vykreslete bodový diagram závislosti obsahu alkoholu na hustotě vína a zakreslete též odhadnutou regresní přímku. Obrázek vložte do textového dokumentu.
6. Ověřte (graficky i testem) všechny předpoklady regresního modelu. Obrázky vložte do textového dokumentu a okomentujte.
7. Pomocí koeficientu determinace ohodnoťte kvalitu svého modelu: napište jeho číselnou hodnotu a interpretujte ji.